Supplementary Information for

The utility of native MS for understanding the mechanism of action of repurposed therapeutics in COVID-19: heparin as a disruptor of the SARS-CoV-2 interaction with its host cell receptor.

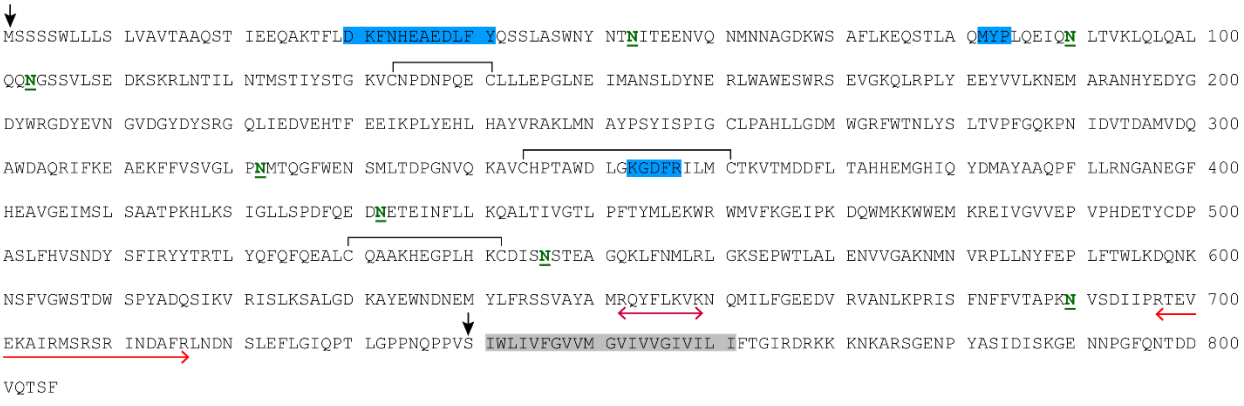
Yang Yang, Yi Du and Igor A. Kaltashov*

Department of Chemistry, University of Massachusetts-Amherst, 240 Thatcher Way, Amherst, MA 01003

Contents

Scheme S1: ACE2 amino acid sequence

The sequence is based on the UniProt entry Q9BYF1. The black vertical arrows indicate the termini of the construct used in this work (not inclusive of the C-terminal His-tag). The SARS-CoV-2 S-protein RBD contact sites are highlighted in blue. The presumed N-glycosylation sites are underlined, and the disulfide bonds are indicated with brackets. The gray-shaded region (not included in the construct) is the transmembrane part of ACE2.



Scheme S2: RBD amino acid sequence

The RBD sequence (underlined with blue lines) is shown in the context of UniProt entry P0DTC2 representing the entire ectodomain of the SARS-CoV-2 S-protein (the gray-shaded region corresponds to the S1 domain). The amino acid residues forming distinct positive patches on the RBD surface are colored in blue. The presumed N-glycosylation sites are underlined, and the N- and O-glycosylation sites reported by Shajahan, et al. (Deducing the N- and O- glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology* **2020**, in press**)** are indicated with the blue and yellow boxes, respectively (the blue crosses indicate the presumed glycosylation sites that were reported to be glycan-free by Shajahan, et al.). The disulfide bonds are indicated with black brackets. The red-shaded regions indicate the protease cleavage sites. The fusion peptide and the heptad repeats are underlined with green and brown lines, respectively.

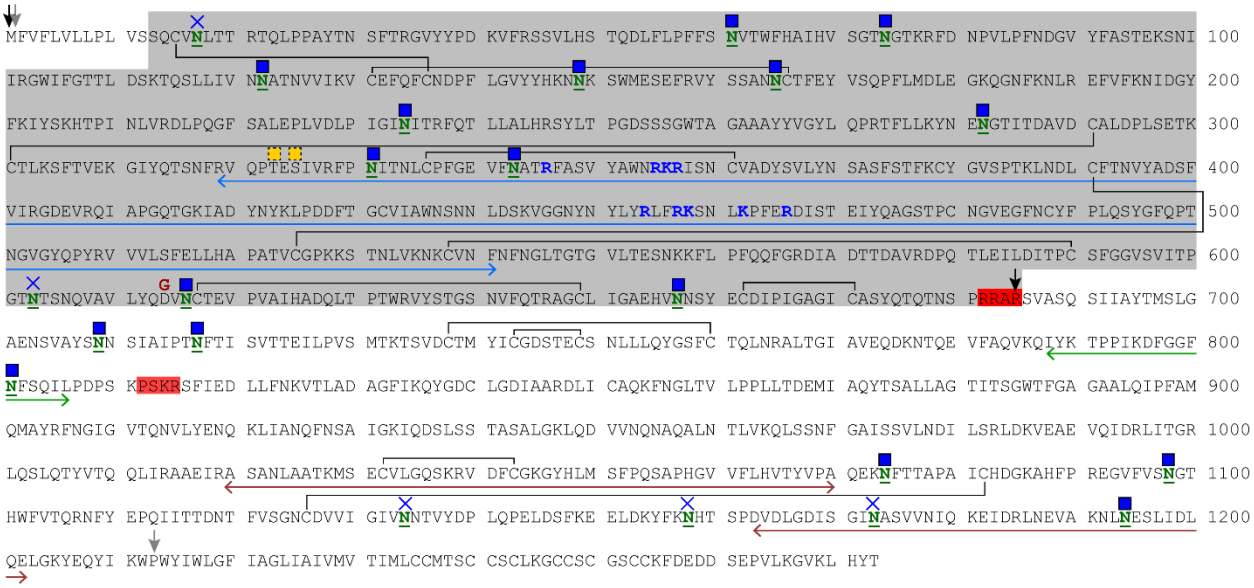


Figure S1. Intact mass analysis of RBD expressed in *E. coli* (RayBiotech)

An ESI mass spectrum of an aqueous solution of RBD expressed in a prokaryotic system (solvent composition: 50 mM Na_2HPO_4 at pH 7.4, 0.5 M NaCl, 450 mM imidazole, and 6 M urea). The discernable charge state ladder in the *m/z* region 1,000-2,000 (see the inset for a zoomed view) corresponds to ionic species with a mass of 25±1 kDa (better precision could not be achieved due to the extensive adduct formation). The abundant peaks in the low *m/z* region (below 1,000) correspond to $[Na(NaCl)_n]^+$ and $[Na(NaCl)_n]^{2+}$ clusters. The unresolved spectral features at high *m/z* region (above 2,000) likely represent covalently linked (via intermolecular disulfide bonds) aggregates of RBD.

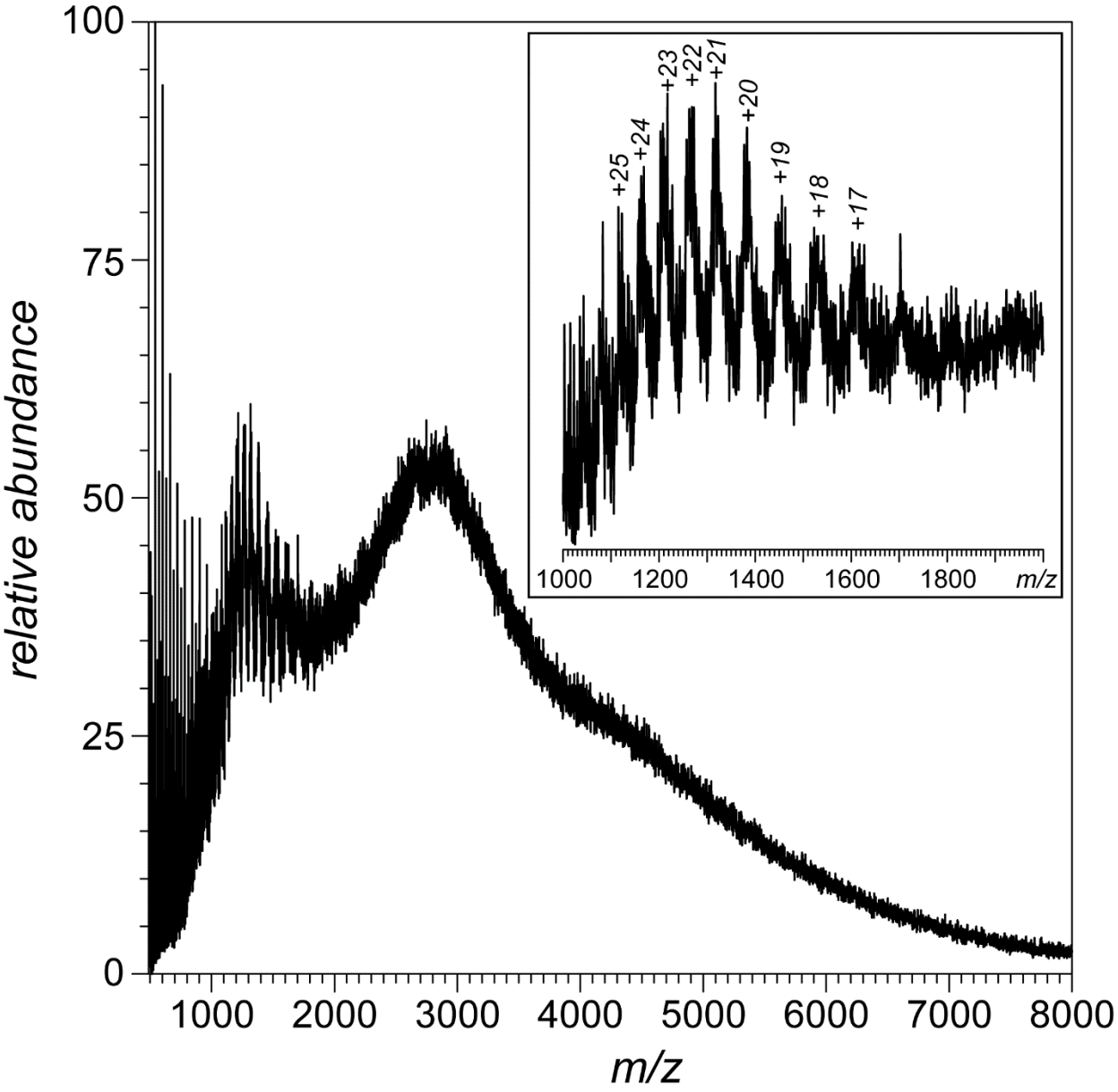


Figure S2. Intact-mass analysis of RBD expressed in a baculovirus/insect cells system (Sino Biological)

Black trace: an ESI mass spectrum of RBD expressed in a eukaryotic system (solvent conditions: $H_2O/CH_3CN/HCO_2H$, 50:49.9:0.1 v:v:v). Deconvoluted mass corresponding to the ionic species giving rise to partially resolved peaks in *m/z* region 2,000-3,5000 is 32.3±0.5 kDa; the mass of the ionic species at higher m/z (3,500-5,000) was determined using limited charge reduction (red trace), yielding 64.5 kDa. The gray trace is a reference spectrum acquired under near-native conditions (see Figure 1 of the main text). The presence of the dimeric RBD species under both near-native and denaturing conditions suggests that the two polypeptide chains within the dimer are linked covalently (most likely via inter-molecular disulfide bridge utilizing unpaired Cys^{538} side chain within each polypeptide chain)

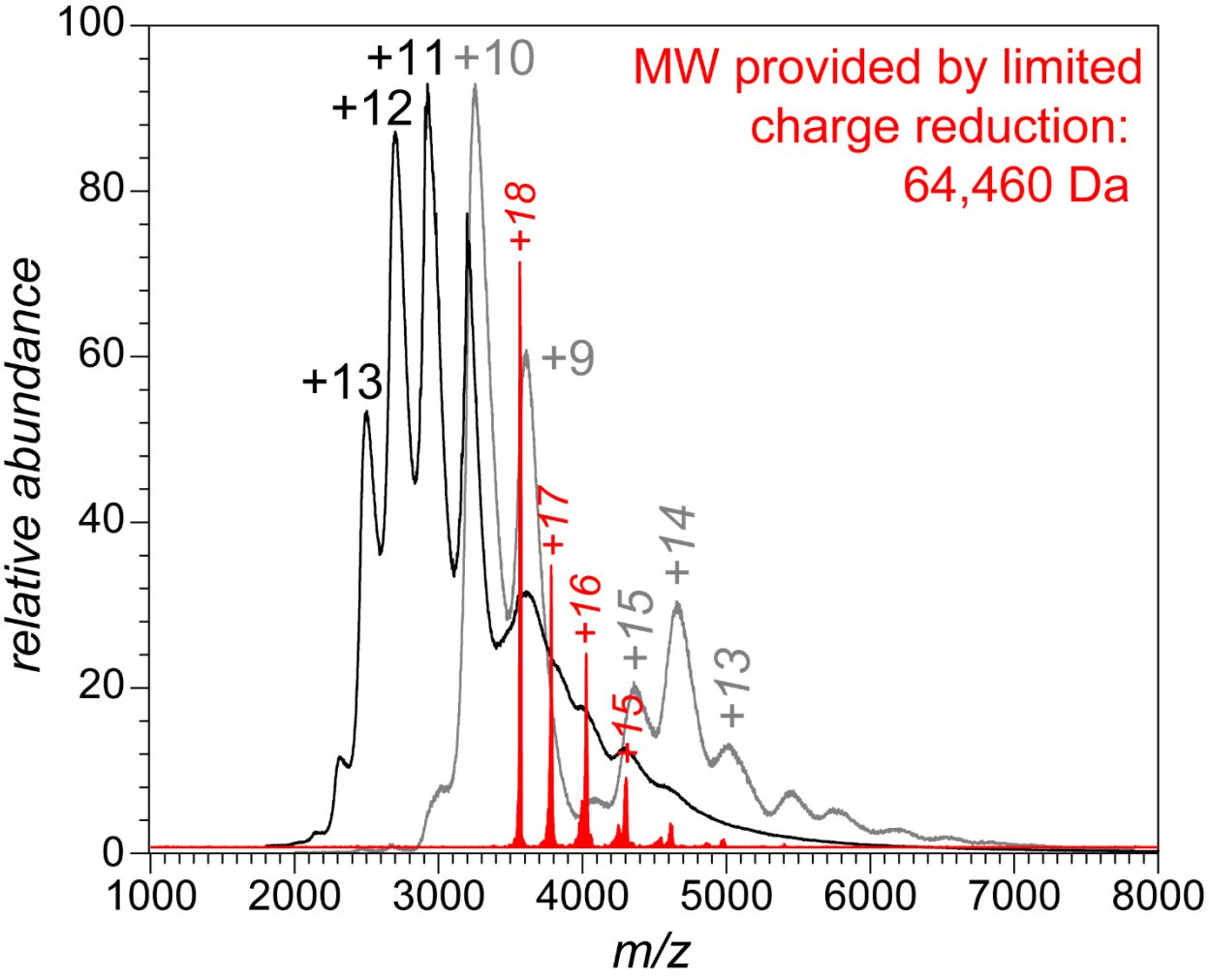


Figure S3. Native MS of ACE2

Black trace: a mass spectrum of an aqueous solution of ACE2 ectodomain (5 µM in 150 mM $NH_4CH_3CO_2$, pH 6.9). Green, orange and red traces: signal generated by limited charge reduction of ionic populations selected at *m/z* 5,161 u, 6,600 u and 6,855 u, respectively (the masses deduced from these charge ladders are 92,954 Da and 185,264 Da, representing the monomeric and dimeric states of ACE2 ectodomain)

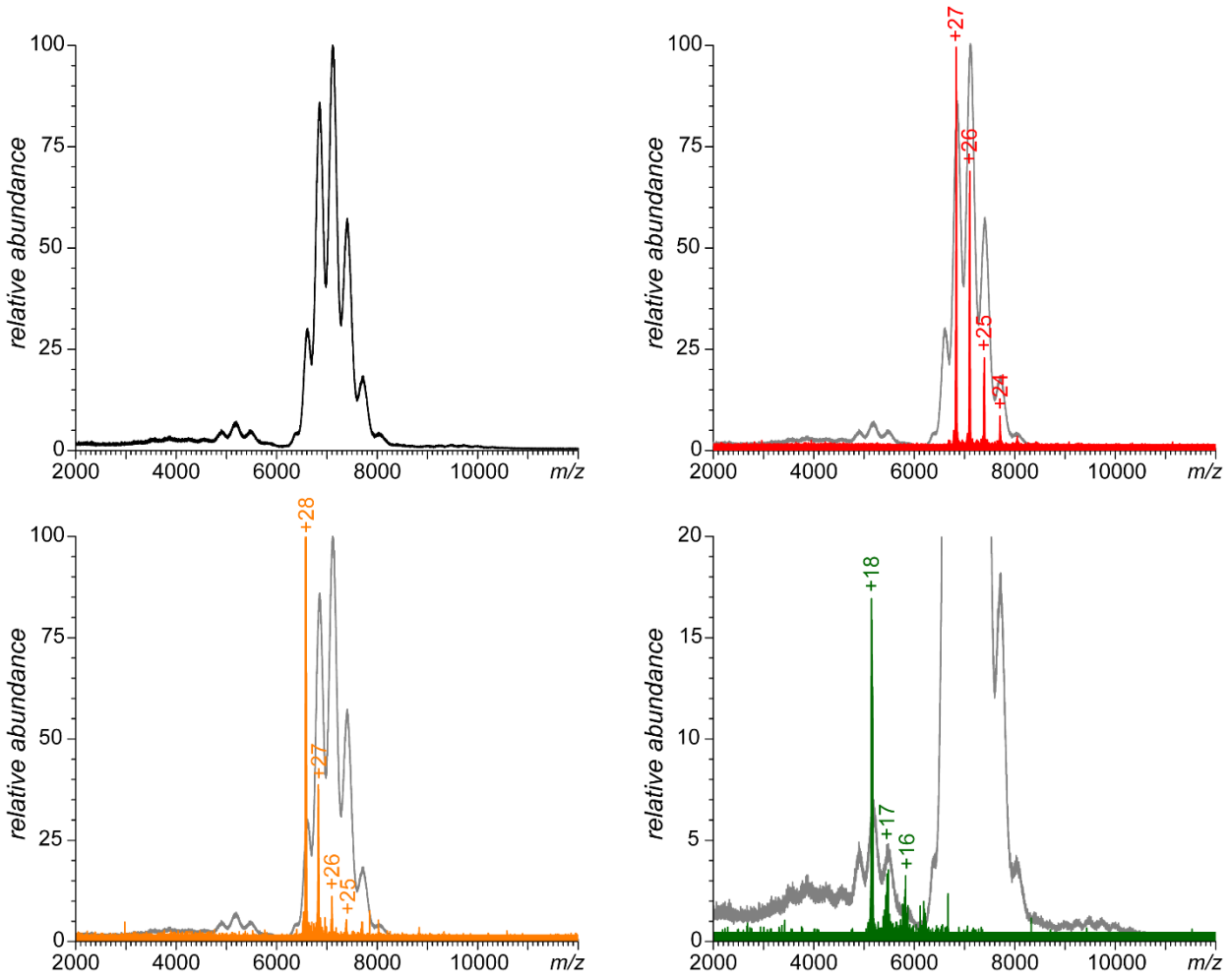


Figure S4. Conformational changes within the ACE2-binding interface region of RBD revealed MD simulations of short heparin oligomer binding.

The ribbon diagrams show backbones of RBD bound to ACE2 (green) and associated with fondaparinux following 1.5 ns MD simulation (teal). Side chains of all contact residues are shown in stick representation; the five critical residues are colored in each structure (red and purple, respectively).

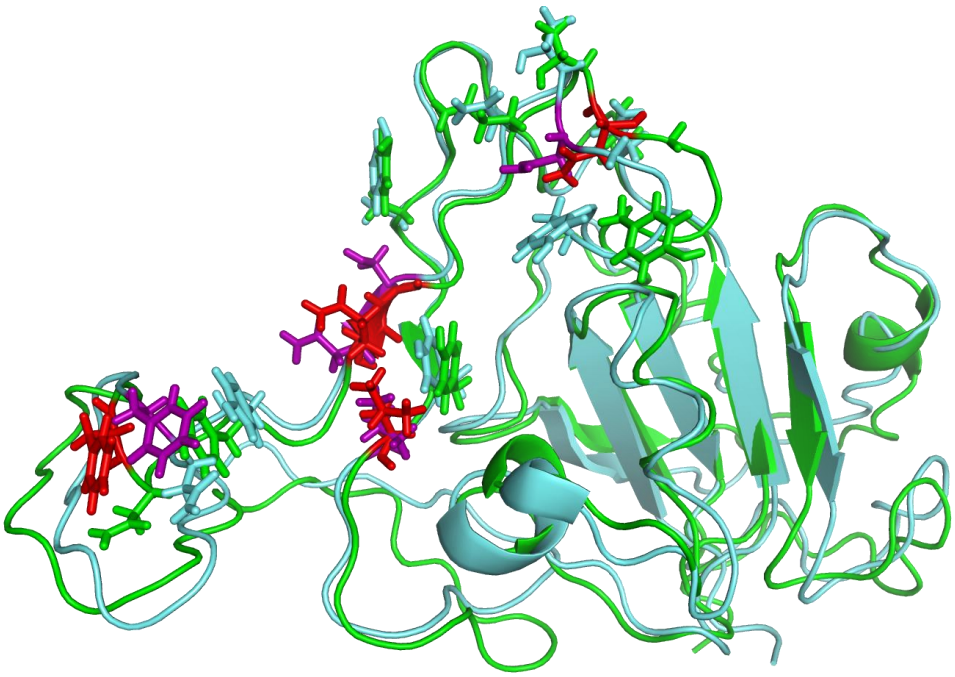


Figure S5. Binding energy evolution for the RBD/dp20 complex over the 6 ns MD simulation window.

Left: evolution of the Coulombic energy and the total free energy of the RBD/dp20 complex (red line; energy evolution for the RBD/fondaparinux complex is shown with gray lines for comparison). Five representative structures of dp20 chains (uniformly colored and shown in stick representation) collected over the 6 ns MD simulation run. The heparin oligomer chains shown in this figure represent structures collected at frame 50 (pink), 200 (black), 500 (orange), 750 (gray) and 1000 (green). The semi-transparent isopotential surfaces show the electrostatic field distribution around the protein (heparin contributions are not taken into account) and correspond to +3*k*T/*e* (blue) and -3*k*T/*e* (red).